# Requesting GPU Resources

There are currently two types of GPU nodes available: NVIDIA Kepler K40 GPU nodes, and NVIDIA Volta V100 GPU nodes. This article describes how to request each type for your PBS job.

## Requesting K40 GPU Nodes (san_gpu)

In your PBS script or **qsub** command line, specify the **san_gpu** model type and **k40** queue name. Each **san_gpu** node is treated as one unit for exclusive access by a single job.

```
#PBS -l select=xx:ncpus=yy:model=san_gpu
#PBS -q k40
```

## Requesting V100 GPU Nodes (sky_gpu and cas_gpu)

The method for requesting V100 GPU nodes changed on July 16, 2020. Instead of each node being treated as one unit for exclusive access by a single job, the nodes are now logically split into two vnodes, one for each socket and its associated CPU cores, GPUs, and memory. These resources are now sharable by different jobs. For more detailed information, see Changes to PBS Job Requests for V100 GPU Resources.

Important: To support these changes and new features, the V100 GPU nodes are currently being served by PBS server pbspl4, until further notice. You can either add **#PBS -q *queue_name*@pbspl4** in your PBS script or the **qsub** command line, or you can **ssh** directly from a Pleiades front end (PFE) to pbspl4 to submit a job.

Access to the V100 GPU nodes is provided through the **v100** queue and the **devel** queue. Each is described below.

WARNING: Your requested amount of CPU memory will be enforced. Once your job uses all of your requested amount of memory, it will be terminated.

## v100 Queue

All of the **sky_gpu** and **cas_gpu** nodes that contain four V100 cards are accessed through the **v100** queue at all times.

The two **sky_gpu** nodes that contain eight V100 cards are accessed through the **v100** queue on weekends only; during the week, they are accessed through the devel queue.

To find default resources assigned to a job, run:

```
pfe% qstat -fQ v100@pbspl4 | grep default
```

The following examples will allocate resources in the **sky_gpu** nodes. Change **model=sky_gpu** to **model=cas_gpu** if you want to use the **cas_gpu** nodes instead. Additionally, the maximum number of CPUs you can request for Cascade Lake-based GPUs is 48 (**ncpus=48**) per node, whereas the max for Skylake-based GPUs is 36.

This request will result in 1 CPU, 1 GPU, and 32 GB of CPU memory:

```
#PBS -q v100@pbspl4
with
#PBS -lselect=1:model=sky_gpu
```

or

```
#PBS -lselect=1:model=sky_gpu:ncpus=1:ngpus=1:mem=32GB
```

The minimum number of CPUs that can be allocated to a job is 1. The minimum amount of memory is 256 MB.

To request exclusive access to *N* number of nodes, run:

```
#PBS -q v100@pbspl4
#PBS - lselect=N:model=sky_gpu:ncpus=36:ngpus=4:mem=360GB
#PBS -lplace=scatter:exclhost
```

Note: Constraints for the **v100** queue are currently set as:

- Two jobs per user in the running or queued state.
- Maximum of 16 nodes per job.
- Maximum of 24 hours walltime.

## devel Queue

For development work between 8:00 a.m Monday and 5:00 p.m. Friday (Pacific Time), the two **sky_gpu** nodes that contain eight V100 cards are reserved for the **devel** queue, with a maximum of two hours and two nodes in a job.

The syntax for requesting resources via the **devel** queue is the same as that described for the **v100** queue, except the queue name is **devel** and the maximum value for **ngpus** is 8.

Reminder: When you run GPU jobs from the **devel** queue, you must use the PBS server pbspl4.

TIP: You can verify the CPUs and/or GPUs assigned to your job by using one of these options:

- ***GPU_hostname*% /u/scicon/tools/bin/clist.x -g**
- ***GPU_hostname*% nvidia-smi**

For detailed instructions describing how to request V100 GPU nodes, see Changes to PBS Job Requests for V100 GPU Resources.

## Checking Job Status

To check the status of your jobs submitted to the **k40** queue, the **v100** queue, or the **devel** queue, run the **qstat** command on the specified host as follows:

```
pfe% qstat k40 -u your_username

pbspl4% qstat v100 -u your_username
or
pfe% qstat v100@pbspl4 -u your_username

pbspl4% qstat devel -u your_username
or
pfe% qstat devel@pbspl4 -u your_username
```

Article ID: 646
Last updated: 14 Oct, 2021

Revision: 43
Running Jobs with PBS -> PBS on GPU Nodes -> Requesting GPU Resources
https://www.nas.nasa.gov/hecc/support/kb/entry/646/